Extended Abstracts of the 1992 International Conference on Solid State Devices and Materials, Tsukuba, 1992, pp. 481-483

Invited

Fabrication and Measurements of Ultra-Short Silicon MOSFETs

Allan HARTSTEIN

IBM Research Division
Thomas J. Watson Research Center
Yorktown Heights, New York

Double Al gate silicon MOSFETs have been fabricated with channel lengths ranging from 125 nm to 7.7 nm using a novel step/edge technique. The devices exhibit lateral tunneling between inversion layer source/drain extensions in the sub-threshold region and inversion layer conduction above threshold. At 0.45K and in devices shorter than 21 nm periodic oscillations are observed at bias voltages where the short gate region is inverted. These are attributed to quantum interference arising from the ultra-short channel length. At bias voltages where the short gate region is in depletion, lateral tunneling is observed between the two inversion layer contacts. The lateral tunneling consists of a superposition of resonant tunneling peaks caused by defect states in the short channel region.

We have used a step/edge technique[1] to fabricate silicon MOSFETs with gate lengths as small as 7.7 nm. These are the shortest channel length devices reported. The inset to Fig. 1 shows an illustration of the device structure. It consists of a 2 Ω-cm, (100), p-type silicon substrate with n^+ source and drain implantations. The short Al gate is formed by evaporation on a step in the SiO_2 gate oxide. A layer of low temperature plasma-enhanced chemical-vapor-deposited (PECVD) SiO_2 forms an insulator for a second Al gate overlaying the device. This gate is used to provide inversion layer extensions for the source and drain contacts.

The gate region was made by first forming a 100 nm thermal oxide and then producing a vertical step by a reactive-ion-etching (RIE) process. The oxide under the short gate was regrown to thicknesses of 22.8 nm and 10.4 nm on different wafers. The source and drain implantations were lithographically patterned to be between 1 μm and 8 μm apart for different devices. The short Al gate was deposited at a grazing angle of incidence. This left Al island films on the surface of the oxide. These films were plasma oxidized to form a non-conducting Al_2O_3 film.

The devices were mounted and cooled to low temperatures. A large positive voltage was applied to the upper gate in order to induce the source/drain extension inversion layer contacts. Conductance was then measured as a function of the voltage applied to the short gate. The conductance was measured at 390 Hz with a source/drain voltage of 1 mV. Fig. 1 shows the conductance of six devices at 77 K. The figure shows regions of lateral tunneling for low gate biases (exponential regime) and regions of metallic conductance at high gate biases.

In the metallic regime the contact resistance of the source/drain extensions (including the leads, implantations, inversion layer extensions, and interfaces) is important. This leads to the saturation of the curves in Fig. 1 at high gate voltages. One can correct for this series resistance[1] by noting that the contact resistance is nearly independent of gate voltage, whereas the channel resistance varies linearly with gate voltage. The intrinsic channel conductance (contact resistance subtracted) of these devices is shown in Fig. 2. The ratio of the

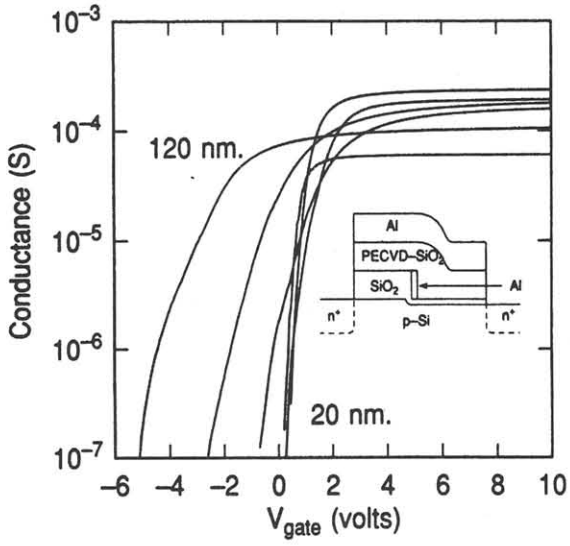

Fig. 1 The conductance is shown as a function of gate voltage for 6 devices with 22.8 nm gate oxides. As indicated, 3 have 20 nm and 3 have 120 nm channel lengths. These data were obtained at 77 K with the upper gate biased at 30 V. The inset shows a schematic section of the short channel MOSFET device. The narrow Al gate ranged from 7.7 nm to 125 nm long. Its gate oxide was either 10.4 nm or 22.8 nm thick.

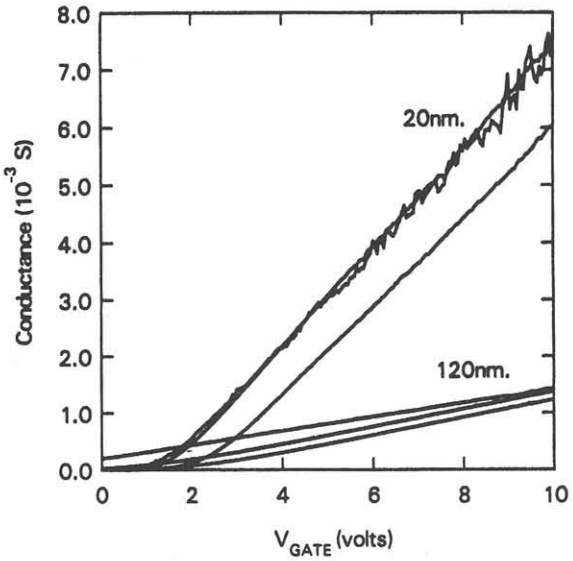

Fig. 2 Intrinsic conductance is shown for the same 6 devices as in Fig. 1. Contact resistance has been subtracted.

slopes for the devices is approximately equal to the ratio of the gate lengths indicating that the effective channel length is equal to the physical gate length[1].

Fig. 3 shows the output I-V curves for a 20 nm device. No correction has been made for the contact resistance. This causes the bunching of the curves at high gate bias.

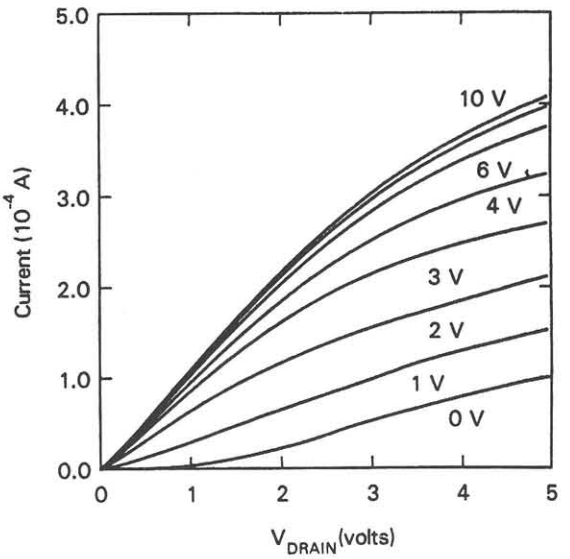

Fig. 3 Current is shown as a function of drain voltage for several values of the applied gate voltage. These data were obtained at 77 K with an upper gate bias of 30 V for a 20 nm device with a 22.8 nm gate oxide.

Fig. 4 shows the channel conductance of four different devices, ranging in length from 7.7 nm to 21 nm, at 0.45K. The experimentally determined contact resistance has been removed. The curves are plotted as a function of gate field rather than gate voltage since two different oxide thicknesses are represented. It is clear that the curves are not smooth and some oscillatory structure is present[2].

In ultra-short channel devices, at low temperatures, one would expect oscillatory structure due to quantum mechanical interference of the electron wave functions within the device. Maxima in the conductance should occur whenever an integral number of half-wavelengths of the electron wave function fit within the device length. The calculated positions of the conductance maxima, using the 2D electron wave functions in the inversion layer, are indicated in Fig. 4 by arrows and associated labels denoting each resonance. It is clear that most of these resonance positions correspond to peaks in the measured conductances. However, two predicted peaks are not seen experimentally. These are peaks at $3\lambda/2$ for the 14 nm and 21 nm long devices.

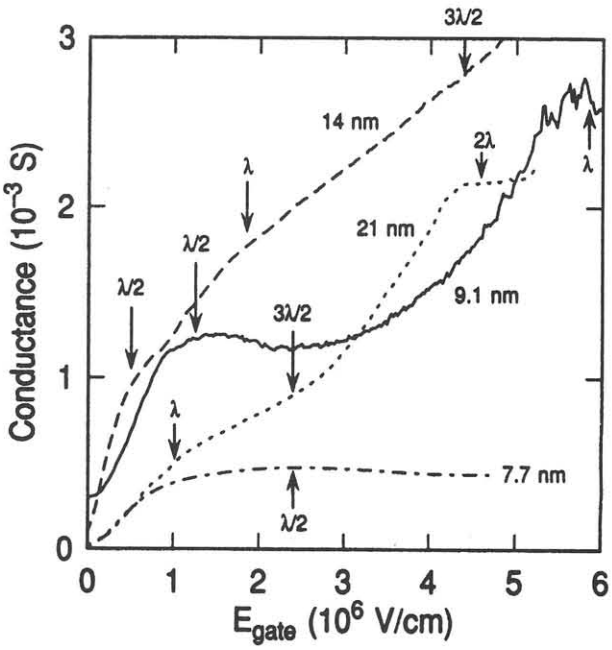

Fig. 4 The conductance is shown as a function of the gate electric field for 4 devices at 0.45 K in the metallic regime. Each curve is for a different gate length as indicated. The arrows and associated labels refer to the predicted positions of conductance maxima for each curve.

A detailed calculation of the quantum mechanical transmission of the source-channel-drain sandwich structure[2] is in good agreement with the data. Some peaks are not seen because the potential discontinuities at the device boundaries are small for certain bias conditions.

Fig. 5 shows a Fowler-Nordheim plot of the drain voltage dependence of the lateral tunneling region of a 7.7 nm device. The high voltage portion of these curves show Fowler-Nordheim tunneling, whereas the low voltage portions display direct tunneling. These results may be modelled using reasonable parameters in a 1D tunneling calculation.

Fig. 6 shows the temperature dependence of the lateral tunneling in a 43.7 nm device. The high temperature region is activated, and the current saturates in the low temperature regime. These curves may also be adequately modelled. In the lateral tunneling regime the

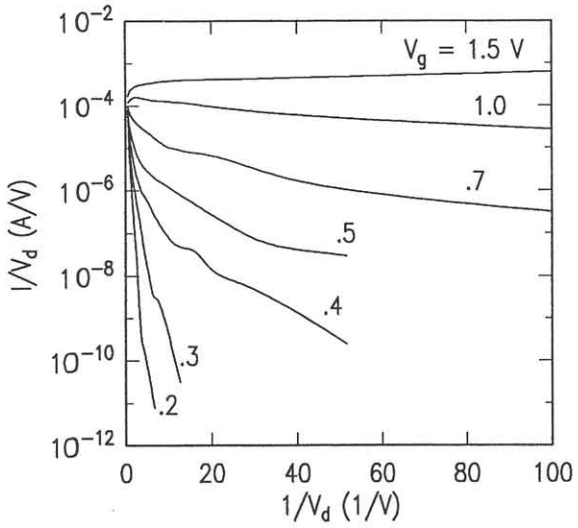

Fig. 5 The drain voltage dependence of the lateral tunneling is shown as a Fowler-Nordheim plot appropriate for a 2D lateral system. These data were obtained at 4.2 K with an upper gate bias of 15 V for a 7.7 nm device.

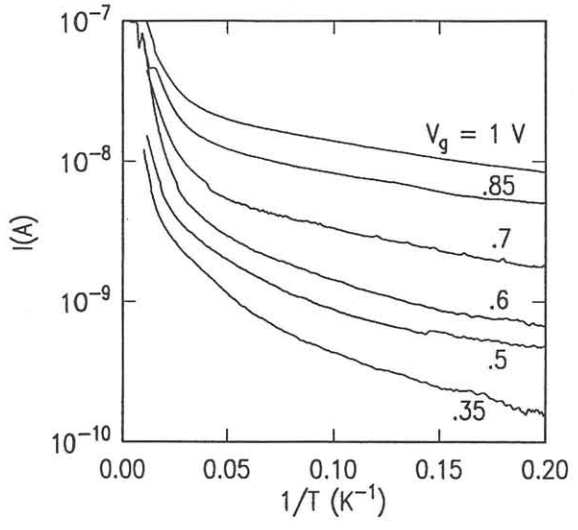

Fig. 6 The temperature dependence of the lateral tunneling in a 43.7 nm device is shown as an activation plot. These data were obtained with 20 V on the upper gate and 1 mV on the drain.

short channel region of the device is biased into depletion and forms a tunneling barrier between the two inversion layer extensions. The current is activated over this barrier at high temperatures and flows by either direct or Fowler-Nordheim tunneling at low temperatures.

A serious problem arises in these model calculations. Different parameters are needed to model each measurement on a single device. The derived barrier heights from each measurement differ by about 30%. It is also difficult to explain the sharp sub-threshold gate voltage dependence of the current shown in Fig. 1.

Fig. 7 shows the gate voltage dependence of the lateral tunneling at 0.45K. Extremely large (3 orders of magnitude) reproducible structure is observed in the direct tunneling curves. The most likely explanation of this is resonant tunneling via single electronic states in the barrier[3]. Similar effects have been observed in tunneling through thin gate oxides[4] and in small MOSFET devices[5,6].

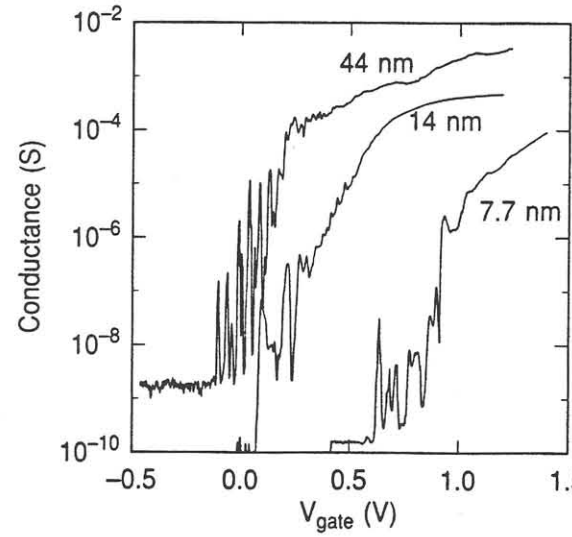

Fig. 7 The conductance is shown as a function of gate voltage for 3 devices in the lateral tunneling regime. These data were obtained at 0.45 K. Each curve is for the device length indicated. The curve for the 44 nm device was shifted up by a factor of 30 for clarity.

If the peaks are indeed due to resonant tunneling, then the lateral tunneling data at higher temperatures, result from a thermally broadened spectrum of the resonant tunneling states. In that case the model calculations used to fit the data in Figs. 5 and 6 are inadequate. The single electron states may be interface states in the short channel region.

In conclusion silicon MOSFET devices have been successfully fabricated with channel lengths as short as 7.7 nm. These devices have been used to investigate quantum mechanical interference caused by the ultra-short gate length. They have also been used to study lateral tunneling between two separate inversion layers.

I wish to thank N. F. Albert, A. A. Bright, S. B. Kaplan, B. Robinson and J. A. Tornello for help in defining the process for fabricating these devices and in their fabrication. I also with to thank N. F. Albert for assistance with some of these measurements.

REFERENCES

1) A. Hartstein, N.F. Albert, A.A. Bright, S.B. Kaplan, B. Robinson and J.A. Tornello, J. Appl. Phys. **68**(1990) 2493.

2) A. Hartstein, Appl. Phys. Lett. **59**(1991) 2028.

3) A. Hartstein, Surf. Sci. **263**(1992) 162.

4) R. Koch and A. Hartstein, Phys. Rev. Lett. **54**(1985) 1848.

5) A.B. Fowler, G.L. Timp, J.J. Wainer, and R.A. Webb, Phys. Rev. Lett. **57**(1986) 138.

6) T.E. Kopley, P.L. McEuen, and R.G. Wheeler, Phys. Rev. Lett. **61**(1988) 1654.